Assessing baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution

Arnošt Komárek[1,*], Emmanuel Lesaffre[1] and Catherine Legrand[2]

[1] *Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium*

[2] *European Organisation for Research and Treatment of Cancer, E. Mounierlaan 83/11, 1200 Brussels, Belgium*

SUMMARY

In multicenter studies, often unknown sources of heterogeneity between centers are present. Moreover, there is not only heterogeneity with respect to the baseline characteristics but also heterogeneity with respect to the efficacy of the treatment. To account for such unknown sources of heterogeneity, we extended the accelerated failure time model with a penalized normal mixture as an error distribution suggested by Komárek and Lesaffre [1] by inclusion of multivariate random effects following a normal

*Correspondence to: arnost.komarek@med.kuleuven.be, Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

Contract/grant sponsor: Research Funds Katholieke Universiteit Leuven; contract/grant number: PDM/06/242

Contract/grant sponsor: Belgian Federal Science Policy Office; contract/grant number: P5/24

Contract/grant sponsor: National Cancer Institute, U.S.A.; contract/grant number: 5U10 CA11488-35

distribution. For computational convenience, we base the inference for the proposed model on the Bayesian methodology with the use of Monte Carlo Markov chain techniques. The proposed method will be illustrated on the disease free survival times of early breast cancer patients collected in the EORTC trial 10854. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: multicenter study; penalized normal mixture; regression; survival analysis

1. INTRODUCTION

The EORTC trial 10854 (Clahsen et al. [2]; van der Hage et al. [3]) is a large multicenter study ($n = 2\,793$ patients in $N = 14$ centers) aiming to compare perioperative polychemotherapy (POP FAC arm) with no further treatment (control arm) on the disease free survival (DFS) time in early breast cancer patients who underwent potentially curative surgery. The centers are located in 5 geographical regions: the Netherlands, Poland, France, Southern Europe, and South Africa. To improve the efficiency with which the treatment effect is evaluated, we want to account for known sources of variability – known patient- and center-specific characteristics (covariates) and use an appropriate regression model. Note that the observed DFS time is often right-censored.

The proportional hazards (PH) model (Cox [4]) is a popular tool to quantify the effect of covariates on the time to event. Let $\boldsymbol{x}_{i,l} = (x_{i,l,1}, \ldots, x_{i,l,s})'$ ($i = 1, \ldots, N$, $l = 1, \ldots, n_i$) denote the covariate vector for the lth patient in the ith center. For the PH model, the hazard function of the event for the (i,l)th patient is expressed by

$$\hbar(t \,|\, \boldsymbol{x}_{i,l}) = \hbar_0(t)\, \exp(\eta_{i,l}), \qquad t > 0, \tag{1}$$

where $\hbar_0$ is an unspecified baseline hazard function, and $\eta_{i,l} = \boldsymbol{\beta}'\boldsymbol{x}_{i,l}$ a linear predictor with

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$ being a vector of regression coefficients. A valuable, although less frequently used, alternative is the accelerated failure time (AFT) model (e.g., Kalbfleisch and Prentice [5], Chap. 7) in which the hazard function $\hbar(t \,|\, \boldsymbol{x}_{i,l})$ is related to the baseline hazard $\hbar_0$ by

$$\hbar(t \,|\, \boldsymbol{x}_{i,l}) = \hbar_0\big\{t \, \exp(-\eta_{i,l})\big\} \, \exp(-\eta_{i,l}), \qquad t > 0. \tag{2}$$

Let $T_{i,l}$ $(i = 1, \ldots, N, \, l = 1, \ldots, n_i)$ denote the event time of the (i,l)th patient. The AFT model (2) can be written in an intuitive way as a simple linear regression model with the logarithmic link function, i.e.

$$\log(T_{i,l}) = \eta_{i,l} + \varepsilon_{i,l}, \tag{3}$$

where $\varepsilon_{i,l}$ are i.i.d. error terms having the distribution of the baseline log-event time. We will assume that the distribution of the error terms is continuous with a density g_ε. Most often, a parametric density g_ε (normal, logistic, Gumbel, ...) is assumed (e.g., Kay and Kinnersley [6]).

1.1. Heterogeneity

In multicenter studies, unknown sources of heterogeneity between centers are often present. This can happen due to many reasons: geographical differences, different working habits of the staff in different centers etc. Moreover, not only the heterogeneity with respect to the baseline characteristics but also the heterogeneity with respect to the efficacy of the treatment may exist. Figure 1 shows Kaplan-Meier estimates of the DFS distribution for the POP FAC arm and the control arm, separately for each center. From these curves, there seems to be heterogeneity among the centers. Not only the overall proportion of DFS patients differs at each time point and in each treatment arm from center to center (*baseline heterogeneity*) but also the effect of treatment on DFS, expressed by the relative position of the two curves in the

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

control and treatment arm seems to vary across centra both quantitatively and qualitatively (*treatment effect heterogeneity*).

<Figure 1 about here.>

The two classical tools which take into account the baseline heterogeneity are the stratified model (e.g., Kalbfleisch and Prentice [5], Sec. 4.4) and a model with the center indicator as one of the covariates (fixed effects model). Similarly, to account for the treatment effect heterogeneity, one can (a) use stratification with respect to the center with treatment interaction; (b) include the center with treatment interaction in the covariate vector $\boldsymbol{x}_{i,l}$. A disadvantage of the first approach is that no direct estimate of the treatment effect is produced. On the other hand, it is debatable whether the results of the fixed effects model can be generalized to a wider population of patients. See also Glidden and Vittinghoff [7] for a discussion to this point. The third, nowadays widely used approach to deal with heterogeneity is the random effects model which is also preferred in the Ref. [7].

1.2. Random effects survival models

Random effects models constitute an alternative to the stratified model or to the fixed effects model. To account for both baseline and treatment effect heterogeneity among centers, we would specify, in either of models (1), (2), (3), the linear predictor $\eta_{i,l}$ as

$$\eta_{i,l} = b_{i,1} + b_{i,2}\mathsf{treat}_{i,l} + \boldsymbol{\beta}'\boldsymbol{x}_{i,l}, \tag{4}$$

where $\mathsf{treat}_{i,l}$ is the treatment indicator for the (i,l)th patient and the vector $\boldsymbol{x}_{i,l}$ contains all covariates but the treatment. Further, the bivariate center specific random effects $\boldsymbol{b}_i = (b_{i,1}, b_{i,2})'$ $(i = 1, \ldots, N)$ are assumed to be i.i.d. with a distribution having a (parametric)

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

density g_b and first 2 moments

$$\begin{aligned} \mathrm{E}(b_{i,1}) &= 0, \\ \mathrm{E}(b_{i,2}) &= \gamma, \end{aligned} \qquad \mathrm{var}(\boldsymbol{b}_i) = \mathbb{D} = \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{1,2} & d_{2,2} \end{pmatrix}, \tag{5}$$

where γ is the mean treatment effect and $d_{1,1}$, $d_{2,2}$, $d_{1,2}$ variance components of the random effects distribution.

In the last decade, the PH model with a univariate random effect $b_i \equiv b_{i,1}$, known also as a *frailty PH model* (e.g., Therneau and Grambsch [8], Hougaard [9], Duchateau and Janssen [10]) became more widely used in practice. The distribution of the random effects b_i is usually specified as either normal for b_i or gamma for $\exp(b_i)$. For the EORTC trial 10854, the frailty PH model which can account only for the baseline heterogeneity among centers, was used by Legrand et al. [11] and we return to it in the discussion.

Nevertheless, it is possible to extend the frailty PH model to include also multivariate random effects. For example, Vaida and Xu [12] consider multivariate random effects having a multivariate normal distribution. A Bayesian estimation of the model with bivariate random effects is presented by Legrand et al. [13]. Other applications of the random effects PH model in the context of the multicenter studies can be found in the literature, e.g., Glidden and Vittinghoff [7], Gray [14], Matsuyama et al. [15], Yamaguchi and Ohashi [16], Yamaguchi et al. [17].

However, the random effects PH model may show some deficiencies. Firstly, for most distributions of random effects, the marginal hazard function obtained by integrating the random effects out does not satisfy the PH assumption any more. That is, the regression coefficients have a clear interpretation only conditionally. In the case the marginal effect of covariates is of interest (typically in epidemiology), it is very difficult to get their correct

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

marginal interpretation. Secondly and more importantly, the effect of the covariate depends on the choice of the density g_b of the random effects. Consequently, the estimates of the regression parameters $\boldsymbol{\beta}$ or the treatment effect γ can be highly sensitive towards, a difficult to check, choice of g_b. See Hougaard [9], Chap. 7 for more details.

These drawbacks do not carry over to the random effects AFT model. Indeed, starting from its linear mixed model representation (3), it is easily seen that the meaning of the regression parameters $\boldsymbol{\beta}$ or the treatment effect γ is the same conditionally, given $\boldsymbol{b}_i$ as well as marginally over $\boldsymbol{b}_i$. Indeed, when the random effects are integrated out from the model (3) with the linear predictor (4), we obtain again model (3) with the linear predictor changed to $\eta_{i,l} = \gamma\,\mathsf{treat}_{i,l} + \boldsymbol{\beta}'\boldsymbol{x}_{i,l}$. The error distribution changes to an appropriate convolution of the random effects distribution and the distribution of the error terms in the random effects model. For this reason, we concentrate here on the random effects AFT model and use its linear mixed model representation (3) in the remainder of the paper.

1.3. Baseline survival distribution

Any parametric assumption concerning the baseline survival distribution in the AFT model (3) represented by the density g_ε is very difficult to check with censored data. For this reason, it is our intention to leave g_ε either unspecified or specify it in a flexible way. Pan and Louis [18] and Pan and Connett [19] consider the univariate random effects AFT model and estimate the distribution of the error term by inclusion of a non-parametric Kaplan-Meier estimation step in their estimation procedure.

An alternative route, namely by using of smoothing techniques, was recently taken by Komárek et al. [20] and Komárek and Lesaffre [1]. In both papers, the error density g_ε is

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

expressed as a penalized normal mixture. The model of the former paper does not include the random effects and cannot take heterogeneity into account, however. Univariate random effects allowing to take into account the baseline heterogeneity, but not the treatment effect heterogeneity, are included in the AFT model suggested in the later paper.

For the analysis of the EORTC trial 10854, we modified the method of Komárek and Lesaffre [1] to include multivariate random effects for which a multivariate normal distribution is assumed. This will allow us to consider both the baseline as well as the treatment effect heterogeneity. The reasons why we assume a normal distribution for the random effects and do not smooth it similarly as the distribution of the error term are the following: (i) The number of centers in our application is quite low (14) providing only a low number of (moreover latent) "observations" to estimate the shape of the distribution; (ii) It has been shown in the literature (Keiding et al. [21], Lambert et al. [22]) that the regression parameters which are usually of the primary interest are robust against misspecification of the random effects distribution; (iii) When the interest lies in the marginal characteristics like the hazard or survival functions, a possible misspecification of the random effects distribution is at least partly corrected by the estimation of the error distribution.

The remainder of the paper is organized as follows. Section 2 describes in detail the proposed random effects AFT model. In Section 3, we describe the inferential procedure for suggested model based on the Monte Carlo Markov chain methodology. The analysis of the DFS time in early breast cancer patients is presented in Section 4. We finalize the paper by a discussion in Section 5.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

2. RANDOM EFFECTS AFT MODEL WITH PENALIZED NORMAL MIXTURE AS AN ERROR DISTRIBUTION

Our approach not only allows for right-censored data but also for left- or interval-censored data. Therefore, assume that $T_{i,l}$ $(i = 1, \ldots, N,\, l = 1, \ldots, n_i)$ occurred within an interval of time $\lfloor t^L_{i,l},\, t^U_{i,l} \rfloor$. For an exactly observed event time, $\lfloor t^L_{i,l},\, t^U_{i,l} \rfloor = [t_{i,l},\, t_{i,l}]$, for a right-censored observation, $\lfloor t^L_{i,l},\, t^U_{i,l} \rfloor = (t_{i,l},\, \infty)$. Further assume that observed intervals are the result of an independent noninformative censoring process.

In this paper, we consider the AFT model (3) with the following linear predictor

$$\eta_{i,l} = \boldsymbol{b}_i' \boldsymbol{z}_{i,l} + \boldsymbol{\beta}' \boldsymbol{x}_{i,l}, \quad i = 1, \ldots, N,\; l = 1, \ldots, n_i,$$

where $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,q})'$ are i.i.d. vectors of random effects with a density g_b, which is assumed here to be the density of the multivariate normal distribution with (unknown) mean $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ and (unknown) covariance matrix $\mathbb{D}$. i.e., $g_b(\boldsymbol{b}_i) = \varphi_q(\boldsymbol{b}_i \,|\, \boldsymbol{\gamma},\, \mathbb{D})$. Further, $\boldsymbol{x}_{i,l}$ is a vector of patient- and center-specific covariates assumed to have an *homogeneous* effect across centers. Finally, $\boldsymbol{z}_{i,l} = (z_{i,l,1}, \ldots, z_{i,l,q})'$ is a vector of factors with a varying (*heterogeneous*) effect across centers. For example, to model the baseline and treatment effect heterogeneity between centers we take $\boldsymbol{z}_{i,l} = (1,\ \mathsf{treat}_{i,l})'$. For identifiability reasons, $\gamma_1 = 0$ whenever the baseline heterogeneity between centers is considered. i.e. whenever $z_{i,l,1} \equiv 1$. For convenience in the notation, we will assume in the remainder of the paper that $z_{i,l,1} \equiv 1$.

To allow for a flexible specification of the baseline survival distribution, represented by the density g_ε of the error terms $\varepsilon_{i,l}$ in the AFT model (3), it is specified as a shifted and scaled penalized normal mixture (see Komárek et al. [20], and Komárek and Lesaffre [1]). That is,

$$g_\varepsilon(\varepsilon) = \tau^{-1} \sum_{j=-K}^{K} w_j(\boldsymbol{a})\, \varphi_1\big\{\tau^{-1}(\varepsilon - \alpha) \,|\, \mu_j,\, \sigma^2\big\}, \tag{6}$$

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

where α and τ are (unknown) intercept and scale parameter, respectively, and

$$w_j(\boldsymbol{a}) = \frac{\exp(a_j)}{\sum_{k=-K}^{K} \exp(a_k)}, \quad j = -K, \ldots, K \tag{7}$$

are (unknown) mixture weights. The weights in (7) are reparametrized to ensure that g_ε is a density for which we need $0 < w_j < 1$, $j = -K, \ldots, K$ and $\sum_j w_j = 1$. Therefore, we will work with the parameter vector $\boldsymbol{a} = (a_{-K}, \ldots, a_K)'$ instead of the vector $\boldsymbol{w} = (w_{-K}, \ldots, w_K)'$. Further, $\boldsymbol{\mu} = \{\mu_{-K}, \ldots, \mu_K\}$ is a *fine* grid of equidistant knots centered around zero ($\mu_0 = 0$) and σ^2 is a fixed basis variance. The following choice, also used in the analysis presented in Section 4, is: $K = 15$, $\mu_{-K} = -4.5$, $\mu_K = 4.5$, $\sigma = 0.2$ and $\mu_{j+1} - \mu_j = 0.3$, see Komárek et al. [20] for a motivation.

2.1. Penalized likelihood

In the following, we use the convention that $\int_c^c p(t)\, dt = p(c)$. The likelihood contribution of the ith center can then be written as

$$L_i = \int_{\mathbb{R}^q} \left\{ \prod_{l=1}^{n_i} \int_{t_{i,l}^L}^{t_{i,l}^U} p(t \,|\, \boldsymbol{a},\, \alpha,\, \tau,\, \boldsymbol{\beta},\, \boldsymbol{b})\, dt \right\} \varphi_q(\boldsymbol{b} \,|\, \boldsymbol{\gamma},\, \mathbb{D})\, d\boldsymbol{b}, \tag{8}$$

where

$$p(t \,|\, \boldsymbol{a},\, \alpha,\, \tau,\, \boldsymbol{\beta},\, \boldsymbol{b}) = (t\tau)^{-1} \sum_{j=-K}^{K} w_j(\boldsymbol{a}) \varphi_1\left(\frac{\log(t) - \alpha - \boldsymbol{b}'\boldsymbol{z}_{i,l} - \boldsymbol{\beta}'\boldsymbol{x}_{i,l}}{\tau} \,\middle|\, \mu_j,\, \sigma^2 \right). \tag{9}$$

To estimate the unknown parameters, we propose to follow the approach of Komárek et al. [20] and maximize the penalized likelihood

$$L^{penal}(\boldsymbol{\theta}) = \prod_{i=1}^{N} L_i \times \exp\Big\{ -\frac{\lambda}{2} \sum_{j=-K+s}^{K} (\Delta^s a_j)^2 \Big\} \tag{10}$$

with respect to $\boldsymbol{\theta} = \big(\boldsymbol{\beta}', \boldsymbol{\gamma}', \text{vec}(\mathbb{D}),\, \alpha,\, \tau,\, \boldsymbol{a}',\, \lambda\big)'$. In expression (10), Δ^s denotes the sth-order difference operator ($s = 3$ was used in the analysis presented in Section 4). The penalty

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

term, $-\frac{\lambda}{2}\sum_{j=-K+s}^{K}(\Delta^s a_j)^2$, which can also be written as $-\frac{\lambda}{2}\,\boldsymbol{a}'\mathbb{P}_s'\mathbb{P}_s\boldsymbol{a}$ for an appropriate difference operator matrix $\mathbb{P}_s$, avoids identifiability problems or overfitting the data, see Eilers and Marx [23]. A trade-off between the smoothness of the density g_ε and fitting the data is driven by the smoothing parameter λ, which has to be estimated as well.

2.2. Bayesian specification

Wahba [24] pointed out the link between penalized likelihood and a Bayesian specification of the model. This link is exploited by Komárek and Lesaffre [1] to obtain estimates of the parameter $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_{-\boldsymbol{a}} = \big(\boldsymbol{\beta}', \boldsymbol{\gamma}', \text{vec}(\mathbb{D}), \alpha, \tau, \lambda\big)'$ and suppose the prior distribution of $\boldsymbol{\theta}_{-\boldsymbol{a}}$ is proportional to a constant (noninformative prior). Suppose further that the prior of the vector $\boldsymbol{a}$ is specified to be a Gaussian Markov random field (GMRF, see, e.g., Besag et al. [25]), namely

$$p(\boldsymbol{a}\,|\,\lambda) \propto \exp\Big(-\frac{\lambda}{2}\,\boldsymbol{a}'\mathbb{P}_s'\mathbb{P}_s\boldsymbol{a}\Big), \tag{11}$$

$$p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \propto 1. \tag{12}$$

Let

$$p(\boldsymbol{\theta}) = p(\boldsymbol{a}\,|\,\lambda) \;\times\; p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \tag{13}$$

be the joint prior distribution of $\boldsymbol{\theta}$. Then, using the Bayes' rule

$$p(\boldsymbol{\theta}\,|\,\text{data}) \;\propto\; \prod_{i=1}^{N} L_i \;\times\; p(\boldsymbol{\theta}) \;=\; \prod_{i=1}^{N} L_i \;\times\; p(\boldsymbol{a}\,|\,\lambda) \;\times\; p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \;\propto\; L^{penal}(\boldsymbol{\theta}). \tag{14}$$

So that, the posterior density of $\boldsymbol{\theta}$ is proportional to the penalized likelihood (10).

Instead of (a difficult) maximization of the penalized likelihood (10), one can infer on the components of $\boldsymbol{\theta}$ from suitable (marginal) characteristics of the posterior distribution $p(\boldsymbol{\theta}\,|\,\text{data})$. The inference can relatively easily be based on a sample from the posterior

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

distribution obtained using MCMC methodology (e.g., Robert and Casella [26]). This approach will be followed here.

2.3. Prior distributions

Note that in our context, it is not possible to be fully noninformative about $\boldsymbol{\theta}_{-\boldsymbol{a}}$ since otherwise the posterior distribution is improper. Hence, we cannot use $p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \propto 1$. Instead, we specify $p(\boldsymbol{\theta}_{-\boldsymbol{a}})$ as a product of vague, but proper distributions. This will ensure that the resulting posterior distribution is proper. Namely,

$$p(\boldsymbol{\theta}_{-\boldsymbol{a}}) = \prod_{j=1}^{s} p(\beta_j) \times \prod_{j=2}^{q} p(\gamma_j) \times p(\mathbb{D}) \times p(\alpha) \times p(\tau^{-2}) \times p(\lambda), \tag{15}$$

where $p(\beta_j)$ $(j = 1, \ldots, s)$, $p(\gamma_j)$ $(j = 2, \ldots, q)$, $p(\alpha)$ are densities of the normal distribution with (zero) mean and large variance, e.g., $\mathcal{N}(0, 10^2)$ was used in the analysis of Section 4. Further, $p(\mathbb{D})$ is the inverse Wishart distribution with a small number of degrees of freedom df_b and a diagonal scale matrix $\mathbb{S}_b$ with small values on the diagonal. In Section 4, we used $df_b = q = 2$ and $\mathbb{S}_b = \text{diag}(0.002)$. Finally, $p(\tau^{-2})$ and $p(\lambda)$, prior densities of the parameters that can be interpreted as inverse variances, are densities of a dispersed gamma distribution, e.g., Gamma(1, 0.005) distributions were used in Section 4.

3. INFERENCE

As mentioned above, we will base the inference on the sample from the posterior distribution (14) which is proportional to the product of (8), (11) and (15).

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

3.1. Bayesian data augmentation

A convenient way to avoid integration over the random effects $\{\boldsymbol{b}_i : i = 1, \ldots, N\}$ and the censored times $\{t_{i,l} : t^L_{i,l} < t^U_{i,l}, i = 1, \ldots, N, l = 1, \ldots, n_i\}$ in the likelihood (8) is to use Bayesian data augmentation (Tanner and Wong [27]).

Further, it is not necessary to work explicitly with the normal mixture (9). Intrinsicly, we can assume that the residual log-event times belong to one of the $2K + 1$ normal components, labeled by $-K, \ldots, K$. Let $r_{i,l}$ $(i = 1, \ldots, N, l = 1, \ldots, n_i)$ be the label of the component to which the (i, l)th residual log-event time belongs, i.e. $\mathrm{P}(r_{i,l} = j \,|\, \boldsymbol{a}) = w_j(\boldsymbol{a})$ $(j = -K, \ldots, K)$.

For convenience we explain the Bayesian data augmentation approach to the case when all event times $t_{i,l}$ are censored, i.e. [data] $= (t^L_{1,1}, t^U_{1,1}, \ldots, t^L_{N,n_N}, t^U_{N,n_N})'$ and $t^L_{i,l} < t^U_{i,l}$ for all i and l. Let $\boldsymbol{\psi} = (\boldsymbol{t}', \boldsymbol{r}', \boldsymbol{B}')'$, with $\boldsymbol{t} = (t_{1,1}, \ldots, t_{N,n_N})'$, $\boldsymbol{r} = (r_{1,1}, \ldots, r_{N,n_N})'$, and $\boldsymbol{B} = (\boldsymbol{b}'_1, \ldots, \boldsymbol{b}'_N)'$, be the vector of latent data, i.e. exact event times, component labels, and random effects, respectively. The posterior distribution (14) can be written as

$$p(\boldsymbol{\theta} \,|\, \text{data}) = \int p(\boldsymbol{\theta}, \boldsymbol{\psi} \,|\, \text{data})\, d\boldsymbol{\psi}. \tag{16}$$

When inference is based on the marginal characteristics of the distribution $p(\boldsymbol{\theta} \,|\, \text{data})$, we can sample from $p(\boldsymbol{\theta}, \boldsymbol{\psi} \,|\, \text{data})$ and ignore the components of $\boldsymbol{\psi}$ in the sample. The distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi} \,|\, \text{data})$ has a relatively simple expression. Indeed, using the Bayes' formula

$$p(\boldsymbol{\theta}, \boldsymbol{\psi} \,|\, \text{data}) \propto p(\text{data} \,|\, \boldsymbol{\theta}, \boldsymbol{t}, \boldsymbol{r}, \boldsymbol{B}) \times p(\boldsymbol{t} \,|\, \boldsymbol{r}, \boldsymbol{B}, \boldsymbol{\theta}) \times p(\boldsymbol{r} \,|\, \boldsymbol{B}, \boldsymbol{\theta}) \times p(\boldsymbol{B} \,|\, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}), \tag{17}$$

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

where

$$p(\text{data} \,|\, \boldsymbol{\theta}, \boldsymbol{t}, \boldsymbol{B}) = p(\text{data} \,|\, \boldsymbol{t}) \propto \prod_{i=1}^{N} \prod_{l=1}^{n_i} I\big\{t_{i,l} \in \lfloor t_{i,l}^L, t_{i,l}^U \rfloor\big\}, \tag{18}$$

$$p(\boldsymbol{t} \,|\, \boldsymbol{r}, \boldsymbol{B}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{l=1}^{n_i} \left\{ (t_{i,l}\tau)^{-1} \varphi_1 \left(\frac{\log(t_{i,l}) - \alpha - \boldsymbol{b}_i' \boldsymbol{z}_{i,l} - \boldsymbol{\beta}' \boldsymbol{x}_{i,l}}{\tau} \,\middle|\, \mu_{r_{i,l}}, \sigma^2 \right) \right\}, \tag{19}$$

$$p(\boldsymbol{r} \,|\, \boldsymbol{B}, \boldsymbol{\theta}) = p(\boldsymbol{r} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{l=1}^{n_i} w_{r_{i,l}}(\boldsymbol{a}), \tag{20}$$

$$p(\boldsymbol{B} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{N} \varphi_q(\boldsymbol{b}_i \,|\, \boldsymbol{\gamma}, \mathbb{D}), \tag{21}$$

and $p(\boldsymbol{\theta})$ is given as a product of (11) and (15). Note that the product of (19) – (21) is in fact equal to the likelihood if the latent data had been observed.

3.2. Markov chain Monte Carlo

To sample from the posterior distribution using the MCMC methodology, we used the Gibbs algorithm (Geman and Geman [28]). The majority of the full conditional distributions are identical to those given by Komárek and Lesaffre [1] and we refer the reader therein. The remaining full conditional distributions pertain to the random effects $\boldsymbol{b}_i$ ($i = 1, \ldots, N$), the means of random effects $\boldsymbol{\gamma}$ and the covariance matrix $\mathbb{D}$ of the random effects. However, they either have a multivariate normal or or an inverse-Wishart distribution. Details are given in the Appendix A.

An R (R Development Core Team [29]) package bayesSurv, freely available from the *Comprehensive R Archive Network* on `http://www.R-project.org`, has been written to sample from the posterior distribution of the model parameters (function bayessurvreg2) and draw the inference (e.g., function predictive2). We illustrate its use in Appendix B.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

3.3. Inference on the model parameters

For each component of the parameter vector $\boldsymbol{\theta}$ we derive summary statistics of the posterior distribution $p(\boldsymbol{\theta} \,|\, \text{data})$, obtained from the MCMC sample, $\boldsymbol{\theta}^{(m)}$ $(m = 1, \ldots, M)$. For example, the posterior median values are approximated by the MCMC sample medians. Highest posterior density (HPD) intervals are derived to express the uncertainty with which the parameter is estimated.

To draw inference on the transformed parameter (vector) $\psi(\boldsymbol{\theta})$, we use the posterior distribution $p\big\{\psi(\boldsymbol{\theta}) \,\big|\, \text{data}\big\}$ and the corresponding MCMC sample $\psi\big(\boldsymbol{\theta}^{(m)}\big)$ $(m = 1, \ldots, M)$. For example, in the context of the AFT model, rather than reporting the results for the fixed effects $\beta_1, \ldots, \beta_s$ or the means $\gamma_2, \ldots, \gamma_q$ of the random effects, we prefer reporting of the acceleration factors $e^{\beta_1}, \ldots, e^{\beta_s}$, or $e^{\gamma_2}, \ldots, e^{\gamma_q}$, respectively. Indeed, they directly determine, how the change in the covariate value accelerates ($e^{\beta} < 1$) or decelerates ($e^{\beta} > 1$) the reference event time.

3.4. Inference on the survival distribution

When interest lies in the survival distribution for a specific combination of covariates $\boldsymbol{x}_{pred}$ and $\boldsymbol{z}_{pred}$, we can compute the predictive survival function $S(t \,|\, \text{data}, \boldsymbol{x}_{pred}, \boldsymbol{z}_{pred})$, or the predictive hazard function $\hbar(t \,|\, \text{data}, \boldsymbol{x}_{pred}, \boldsymbol{z}_{pred})$ $(t > 0)$ from the MCMC output. The procedure is analogous, with only an obvious change in notation, to that described in Komárek and Lesaffre [1] (Section 5.3) and the reader is referred therein for details.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

3.5. Inference on random effects

When dealing with heterogeneity, one might be interested in investigating and explaining the heterogeneity. To this end, we can use the (marginal) posterior distribution $p(\boldsymbol{B} \,|\, \text{data})$ of the random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$, which is obtained from the joint posterior distribution (17) by integrating out the remaining parameters. When an MCMC sample from the joint posterior distribution is available, integration is achieved by simply ignoring these remaining parameters in the sample.

4. THE ANALYSIS OF THE DFS TIME IN EARLY BREAST CANCER PATIENTS

For the analysis of the DFS time in early breast cancer patients in the EORTC trial 10854, we fitted two random effects AFT models, i.e. given by expressions (3) and (4). In both models, we included the following covariates: age group (*<40, 40 – 50, >50 years*), type of prior surgery (*mastectomy, breast conserving*), tumor size (*not palpable or <2 cm, ≥2 cm*), axillary nodal status (*negative, positive*), presence of other related disease (*no, yes*). The first AFT model (**Model with region**) contained also dummies for a geographical location, whereas in the second AFT model (**Model without region**), the geographical location was not included in the covariate vector for fixed effects. Since centers are nested within geographical regions it should be possible to reveal, at least partially, the regional structure of the centers from the estimates of the center-specific random effects $b_{1,1}, \ldots, b_{N,1}$ in the model without region.

For inference we sampled a chain of length 125 000 with 1:5 thinning which took about 2.5 hour on a Pentium IV 2 GHz PC with 512 MB RAM. The last 25 000 iterations of the chain were used to derive the summary statistics.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

4.1. Effect of covariates and the survival distribution

Table I shows the posterior summaries for the acceleration factors revealing the effect of considered covariates in both models. It is seen that the DFS time in the *control* arm is approximately 0.86 times shorter than in the *POP FAC* arm. Based on the model with region included, the DFS time for the middle age group *40 – 50 years* is increased by a factor of 1.38 compared to the youngest group *<40 years*. For the patients from the oldest group *>50 years*, the DFS time is increased by a factor of 1.33 compared to the youngest group. The *breast conserving surgery* increases the DFS time by a factor of 1.26 compared to *mastectomy*. Further, bigger tumors (≥2 cm) lead to a decrease of the DFS time by a factor of 0.63 compared to smaller tumors of size *<2 cm*. A *positive* pathological nodal status decreases the DFS time by a factor of 0.55 compared to a *negative* result. The *presence* of other related disease decreases the PFS time by a factor of 0.72. From the regional effects it is for example seen that *South Africa* performs far the worst than all remaining regions.

<Table I about here.>

In the model without region, the effect of the included covariates is estimated to be practically the same as in the model with region. This illustrates, among other things a general property of the AFT model which is robustness towards omission of important covariates (Hougaard [30]). A complete view on the distribution of the DFS time is given in Figure 2 which shows the predictive hazard and survival functions in the POP FAC and control arm when fixing remaining covariates on their reference values.

<Figure 2 about here.>

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

4.2. Heterogeneity

Figure 3 shows posterior medians and 95% HPD intervals for acceleration factors based on the center-specific random effects in both considered models. For comparison purposes, the plot related to the random intercepts $b_{i,1}$, $(i = 1, \ldots, 14)$ takes also into account the fixed effect of a geographical region in the model with region explicitly included. In the left part of Figure 3, *France* serves as a reference region (model with region) whereas an average over all regions serves as a reference in the right part of Figure 3 (model without region). This causes an overall shift when going from left to right in the upper panel of Figure 3. However, besides that shift, the structure of the posterior medians of the random intercepts is quite similar in both models. That is, the random intercepts in the model without region were able to capture to a large extent the effect of the region.

<Figure 3 about here.>

As one could have expected, omission of the covariate region led to the increase of the variability of the random intercept. Namely, its standard deviation, estimated by the posterior median of $\sqrt{d_{1,1}}$, increased from 0.111 to 0.302, the 95% HPD interval for $\sqrt{d_{1,1}}$ changed from (0.015, 0.292) to (0.142, 0.513).

The lower panel of Figure 3 shows further that treatment effect heterogeneity between centers is of a lower magnitude than the baseline heterogeneity. This is also seen on the posterior medians of the parameter $\sqrt{d_{2,2}}$, standard deviation of $b_{i,2}$ $(i = 1, \ldots, 14)$ which equals to 0.057 in the model with region and to a slightly higher value of 0.074 in the model without region, respectively. The 95% HPD intervals for $\sqrt{d_{2,2}}$ are (0.014, 0.180) and (0.015, 0.212), respectively.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

Most importantly, all increase of the variability caused by the omission of the important covariate (region) was captured by the variance components of the random effects. The residual variability, which has a direct impact on the precision with which the effect of the covariates is evaluated, remains practically the same. More specifically, the posterior median of the standard deviation of the error terms $\varepsilon_{i,l}$ changed from 1.481 in the model with region to 1.470 in the model without region. The corresponding 95% HPD interval changed from (1.341, 1.640) to (1.345, 1.628).

5. DISCUSSION

We have introduced here a possible approach to perform a regression analysis with survival clustered data dealing with a heterogeneity between clusters (centers). Both the baseline heterogeneity, as well as the heterogeneity with respect to the effect of selected covariates has been considered. The heterogeneity has been taken into account by including the random effects in the AFT model. Parametric assumptions concerning the baseline survival distribution have been avoided by using the penalized normal mixture as a model for the error terms in the AFT model.

As we pointed out in Section 1.2, Legrand et al. [11] analyzed the EORTC trial 10854 using the frailty PH model. By considering a *fixed* treatment effect and a *random* center effect their objective was to quantify heterogeneity in outcome over centers. They however do not include a treatment by center interaction and therefore do not account for a possible heterogeneity in the treatment effect between centers. They argue that factor "treatment" cannot explain heterogeneity in outcome found over centers as the same proportion of patients is treated in each treatment arm (randomization stratified by center). However, in our analysis, we

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

demonstrated that the treatment effect heterogeneity cannot be ignored, with a magnitude of this heterogeneity represented by the parameter $\sqrt{d_{2,2}}$. This parameter was, in the model with region, estimated to be half of the magnitude of the baseline heterogeneity represented by the parameter $\sqrt{d_{1,1}}$. For this reason, we do not think that the treatment effect heterogeneity can be automatically ruled out when trying to explain heterogeneity in outcome over centers.

Analogously to Legrand et al. [11], we have found that the baseline heterogeneity between centers is largely explained by the geographical differences. Finally, Legrand et al. [11] compared different centers by the mean of the predicted 5-year DFS rates. A similar comparison was performed in this paper by the mean of the posterior summaries of the acceleration factors based on the center-specific random effects. With respect to the baseline heterogeneity, a similar pattern has been found by both methods.

ACKNOWLEDGEMENTS

The research was primarily supported by Research Grant PDM/06/242, Katholieke Universiteit Leuven. The authors further acknowledge support from the Interuniversity Attraction Poles Program P5/24 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs.

The authors thank the European Organisation for Research and Treatment of Cancer (EORTC) Breast Cancer Group for permission to use the data from EORTC trial 10854 for this research. This research project was further supported by grant number 5U10 CA11488-35 from the US National Cancer Institute (Bethesda, Maryland, USA). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

REFERENCES

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

1. A. Komárek and E. Lesaffre. Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. Technical Report 0546, IAP Statistics Network, Interuniversity Attraction Pole, http://www.stat.ucl.ac.be/IAP, 2005.
2. P. C. Clahsen, C. J. van de Velde, J. P. Julien, J. L. Floiras, T. Delozier, F. Y. Mignolet, and T. M. Sahmoud. Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer. A European Organization for Research and Treatment of Cancer Breast Cancer Cooperative Group Study. *Journal of Clinical Oncology*, 14:745–753, 1996.
3. J. A. van der Hage, C. J. H. van de Velde, J.-P. Julien, J.-L. Floiras, T. Delozier, C. Vandervelden, and L. Duchateau. Improved survival after one course of perioperative chemotherapy in early breast cancer patients: long-term results from the European Organization for Research and Treatment of Cancer (EORTC) Trial 10854. *European Journal of Cancer*, 37:2184–2193, 2001.
4. D. R. Cox. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.
5. J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Chichester, Second edition, 2002.
6. R. Kay and N. Kinnersley. On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. *Drug Information Journal*, 36:571–579, 2002.
7. D. V. Glidden and E. Vittinghoff. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine*, 23:369–388, 2004.
8. T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York, 2000.
9. P. Hougaard. *Analysis of Multivariate Survival Data*. Springer-Verlag, New York, 2000.
10. L. Duchateau and P. Janssen. Understanding heterogeneity in generalized mixed and frailty models. *The American Statistician*, 59:143–146, 2005.
11. C. Legrand, L. Duchateau, R. Sylvester, P. Janssen, J. A. van der Hage, C. J. H. van de Velde, and P. Therasse. Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial. *Clinical Trials*, 3:10–18, 2006.
12. F. Vaida and R. Xu. Proportional hazards model with random effects. *Statistics in Medicine*, 19:3309–3324, 2000.
13. C. Legrand, V. Ducrocq, P. Janssen, R. Sylvester, and L. Duchateau. A Bayesian approach to jointly

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

estimate center and treatment by center heterogeneity in a proportional hazards model. *Statistics in Medicine*, 24:3789–3804, 2005.

14. R. J. Gray. A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, 50:244–253, 1994.
15. Y. Matsuyama, J. Sakamoto, and Y. Ohashi. A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial. *Statistics in Medicine*, 17:1893–1908, 1998.
16. T. Yamaguchi and Y. Ohashi. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statistics in Medicine*, 18:1961–1971, 1999.
17. T. Yamaguchi, Y. Ohashi, and Y. Matsuyama. Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical Methods in Medical Research*, 11:221–236, 2002.
18. W. Pan and T. A. Louis. A linear mixed-effects model for multivariate censored data. *Biometrics*, 56:160–166, 2000.
19. W. Pan and J. E. Connett. A multiple imputation approach to linear regression with clustered censored data. *Lifetime Data Analysis*, 7:111–123, 2001.
20. A. Komárek, E. Lesaffre, and J. F. Hilton. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, 14:726–745, 2005.
21. N. Keiding, P. K. Andersen, and J. P. Klein. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16:215–225, 1997.
22. P. Lambert, D. Collett, A. Kimber, and R. Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, 23:3177–3192, 2004.
23. P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties (with Discussion). *Statistical Science*, 11:89–121, 1996.
24. G. Wahba. Bayesian "confidence intervals" for the cross–validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, 45:133–150, 1983.
25. J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems (with Discussion). *Statistical Science*, 10:3–66, 1995.
26. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, Second edition, 2004.
27. M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal*

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

of the American Statistical Association, 82:528–550, 1987.

28. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayes restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
29. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
30. P. Hougaard. Fundamentals of survival data. *Biometrics*, 55:13–22, 1999.
31. M. Plummer, N. Best, K. Cowles, and K. Vines. *coda: Output analysis and diagnostics for MCMC*, 2005. R package version 0.9-5.

APPENDIX A: MARKOV CHAIN MONTE CARLO

In appendix A, we provide the full conditional distributions for the random effects $\boldsymbol{b}_i$ $(i = 1, \ldots, N)$, the means of random effects $\boldsymbol{\gamma}$ and the covariance matrix $\mathbb{D}$ of the random effects.

Namely,

$$\boldsymbol{b}_i \mid \cdots \sim \mathcal{N}\Big(\mathrm{E}(\boldsymbol{b}_i \mid \cdots), \mathrm{var}(\boldsymbol{b}_i \mid \cdots)\Big), \qquad i = 1, \ldots, N, \tag{22}$$

with

$$\mathrm{E}(\boldsymbol{b}_i \mid \cdots) = \mathrm{var}(\boldsymbol{b}_i \mid \cdots) \times \Big[\mathbb{D}^{-1}\boldsymbol{\gamma} + (\sigma\,\tau)^{-2} \sum_{l=1}^{n_i} \boldsymbol{z}_{i,l}\big\{\log(t_{i,l}) - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_{i,l} - \tau\,\mu_{r_{i,l}}\big\}\Big],$$

$$\mathrm{var}(\boldsymbol{b}_i \mid \cdots) = \Big\{\mathbb{D}^{-1} \;+\; (\sigma\,\tau)^{-2} \sum_{l=1}^{n_i} \boldsymbol{z}_{i,l}\,\boldsymbol{z}'_{i,l}\Big\}^{-1}.$$

Further, let $\boldsymbol{\nu}_{(-1)}$ be the vector of prior means of $\boldsymbol{\gamma}_{(-1)} = (\gamma_2, \ldots, \gamma_q)'$ and $\mathbb{U}_{(-1)}$ be a diagonal matrix having prior variances of $\boldsymbol{\gamma}_{(-1)}$ on the diagonal. Let $\mathbb{V}_{(-1)}$ and $\mathbb{V}_{(-1,1)}$ be the $(2, \ldots, q)$-$(2, \ldots, q)$ block and the $(2, \ldots, q)$-1 block, respectively, of the matrix $\mathbb{D}^{-1}$. Finally, let $\boldsymbol{b}_{i(-1)} = (b_{i,2}, \ldots, b_{i,q})'$ $(i = 1, \ldots, N)$. Then

$$\boldsymbol{\gamma}_{(-1)} \mid \cdots \sim \mathcal{N}\Big(\mathrm{E}(\boldsymbol{\gamma}_{(-1)} \mid \cdots), \mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots)\Big), \tag{23}$$

with

$$\mathrm{E}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) = \mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) \times \Big(\mathbb{U}_{(-1)}^{-1}\boldsymbol{\nu}_{(-1)} + \mathbb{V}_{(-1)} \sum_{i=1}^{N} \boldsymbol{b}_{i(-1)} + \mathbb{V}_{(1,-1)} \sum_{i=1}^{N} b_{i,1}\Big),$$

$$\mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) = \Big(\mathbb{U}_{(-1)}^{-1} \;+\; N\,\mathbb{V}_{(-1)}\Big)^{-1}.$$

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

Finally,

$$\mathbb{D} \,|\, \cdots \sim \text{inverse-Wishart} \left(df_b + N,\ \mathbb{S}_b + \sum_{i=1}^{N} (\boldsymbol{b}_i - \boldsymbol{\gamma})(\boldsymbol{b}_i - \boldsymbol{\gamma})' \right). \tag{24}$$

APPENDIX B: ANALYSIS IN R

Appendix B is devoted to a brief description of the R package bayesSurv to perform the analysis presented in Section 4. We assume that the data are stored in a `data.frame` called `eortc` which has a structure as indicated in Table II. The column `id` identifies patients, column `center` different centers. The DFS time is found in the column `DFStime` and a censoring indicator (0 for right-censored and 1 for observed event times) is given in the column `DFSevent`. The values of covariates are given in columns labeled `trtmt`, `ageGroup`, `typeSur`, `tumSize`, `nodStat`, `otDis`, `region`. The columns corresponding to non-dichotomous covariates (`ageGroup` and `region`) are assumed to be created by the R function `factor` with appropriately chosen reference category.

<Table II about here.>

Firstly, we specify the basis standard deviation σ (`sigma`), the number of knots (`K`), the distance between 2 consecutive knots expressed as a multiple of the basis standard deviation σ (`c4delta`), order of the penalty s (`order`) and prior choices for the intercept α, scale τ and the smoothing hyperparameter λ. Specified choices are stored in lists `params.error` and `prior.error`:

```
> params.error <- list(sigma=0.2)
> prior.error <- list(K=15, c4delta=1.5, order=3,
+   prior.intercept="normal", mean.intercept=0, var.intercept=100,
+   prior.scale="gamma", shape.scale=1, rate.scale=0.005,
+   prior.lambda="gamma", shape.lambda=1, rate.lambda=0.005)
```

Secondly, the prior choices for fixed effects $\boldsymbol{\beta}$, the mean of the random effects b_2 (parameter γ_2) and the random effects $\boldsymbol{b}$ are specified and stored as lists `prior.betaGamma` and `prior.b`.

```
> prior.betaGamma <- list(mean.prior=rep(0, 11), var.prior=rep(100, 11))
```

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

```
> prior.b <- list(prior.D = "inv.wishart", df.D = 2, scale.D = 0.002*c(1,0,1))
```

Note that $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{10})'$ in the model with region so that there are 11 'β' and 'γ' parameters to be estimated.

The core part of the analysis, MCMC sampling, is then performed using the function `bayessurvreg2` in the following way:

```
> library(bayesSurv)
> sample <- bayessurvreg2(Surv(DFStime, DFSevent)~trtmt+ageGroup+typeSur+
+   tumSize+nodStat+otDis+region+cluster(center), random=~trtmt,
+   prior=prior.error, init=params.error,
+   prior.beta=prior.betaGamma, prior.b=prior.b,
+   nsimul=list(niter=125000, nthin=5, nburn=100000), store=list(b=TRUE),
+   dir="/home/userAK/", data=eortc)
```

Sampled chains are then found in the form of ASCII files having an extension `.sim` in the directory called "`/home/userAK/`" and can be further worked out, e.g., using the R package coda [31]. For example, data for Table I were obtained using the following commands:

```
> library(coda)
> betaGamma <- read.table("/home/userAK/beta.sim", header=TRUE)
> exp.betaGamma <- mcmc(exp(betaGamma))
> summary(exp.betaGamma)
> HPDinterval(exp.betaGamma)
```

To compute the predictive hazard and survival functions as shown in Figure 2, we have to specify the combinations of covariates for which the hazard and survival functions would be computed:

```
> eortc.pred <- data.frame(DFStime=c(1, 1), DFSevent=c(0, 0), trtmt=c(1, 0),
+   ageGroup=factor(c(0, 0), levels=0:2, labels=c("<40", "40--50", ">50")),
+   typeSur=c(0, 0), tumSize=c(0, 0), nodStat=c(0, 0), otDis=c(0, 0),
+   region=factor(c(0, 0), levels=0:4, labels=c("F", "NL", "P", "SE", "SA")),
```

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

```
+   center=c(1, 2))
```

Computation of the values of predictive survival and hazard functions on the equidistant grid of 100 time values from 1 to 5 002 days is then performed using the following code:

```
> pred <- predictive2(Surv(DFStime, DFSevent)~trtmt+ageGroup+typeSur+
+   tumSize+nodStat+otDis+region+cluster(center), random=~trtmt,
+   grid=seq(1, 5002, length=100), Gspline=list(dim=1, K=15),
+   quantile=c(0.025, 0.975), only.aver=FALSE, dir="/home/userAK/",
+   predict=list(Surv=TRUE, density=FALSE, hazard=TRUE, cum.hazard=FALSE),
+   data=eortc.pred)
```

By the argument `quantile`, the user can obtain also pointwise posterior predictive quantiles for the hazard and survival function.

More detailed description of the functions from the bayesSurv package and their arguments can be found in the documentation to the package.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

Table I. Posterior medians and 95% highest posterior density intervals for the acceleration factors ($\exp(\gamma)$ and $\exp(\beta)$ parameters).

	Model with region		Model without region	
Effect	Posterior median	95% HPD interval	Posterior median	95% HPD interval
Treatment group (reference: *POP FAC arm*)				
control arm	0.858	(0.712, 1.010)	0.860	(0.729, 1.009)
Age group (reference: *<40 years*)				
40 – 50 years	1.384	(1.035, 1.762)	1.411	(1.064, 1.819)
>50 years	1.330	(1.019, 1.656)	1.368	(1.061, 1.738)
Type prior surgery (reference: *mastectomy*)				
breast conserving	1.257	(1.041, 1.483)	1.281	(1.070, 1.509)
Tumor size (reference: *<2 cm*)				
≥2 cm	0.630	(0.521, 0.748)	0.625	(0.515, 0.745)
Nodal status (reference: *negative*)				
positive	0.549	(0.461, 0.635)	0.546	(0.459, 0.639)
Other disease (reference: *absent*)				
present	0.724	(0.538, 0.930)	0.716	(0.536, 0.926)
Region (reference: *France*)				
The Netherlands	0.669	(0.457, 0.943)		
Poland	1.417	(0.845, 2.154)		
Southern Europe	0.713	(0.465, 1.007)		
South Africa	0.479	(0.295, 0.700)		

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

Table II. Structure of the R `data.frame` `eortc` holding the data.

id	center	DFStime	DFSevent	trtmt	ageGroup
1	11	5 139	0	1	40--50
2	31	4 163	0	0	<40
3	41	733	1	1	>50
⋮	⋮	⋮	⋮	⋮	⋮

typeSur	tumSize	nodStat	otDis	region
0	1	0	0	NL
1	0	0	0	F
0	1	1	0	SE
⋮	⋮	⋮	⋮	⋮

Copyright © 2007 John Wiley & Sons, Ltd.
Prepared using simauth.cls

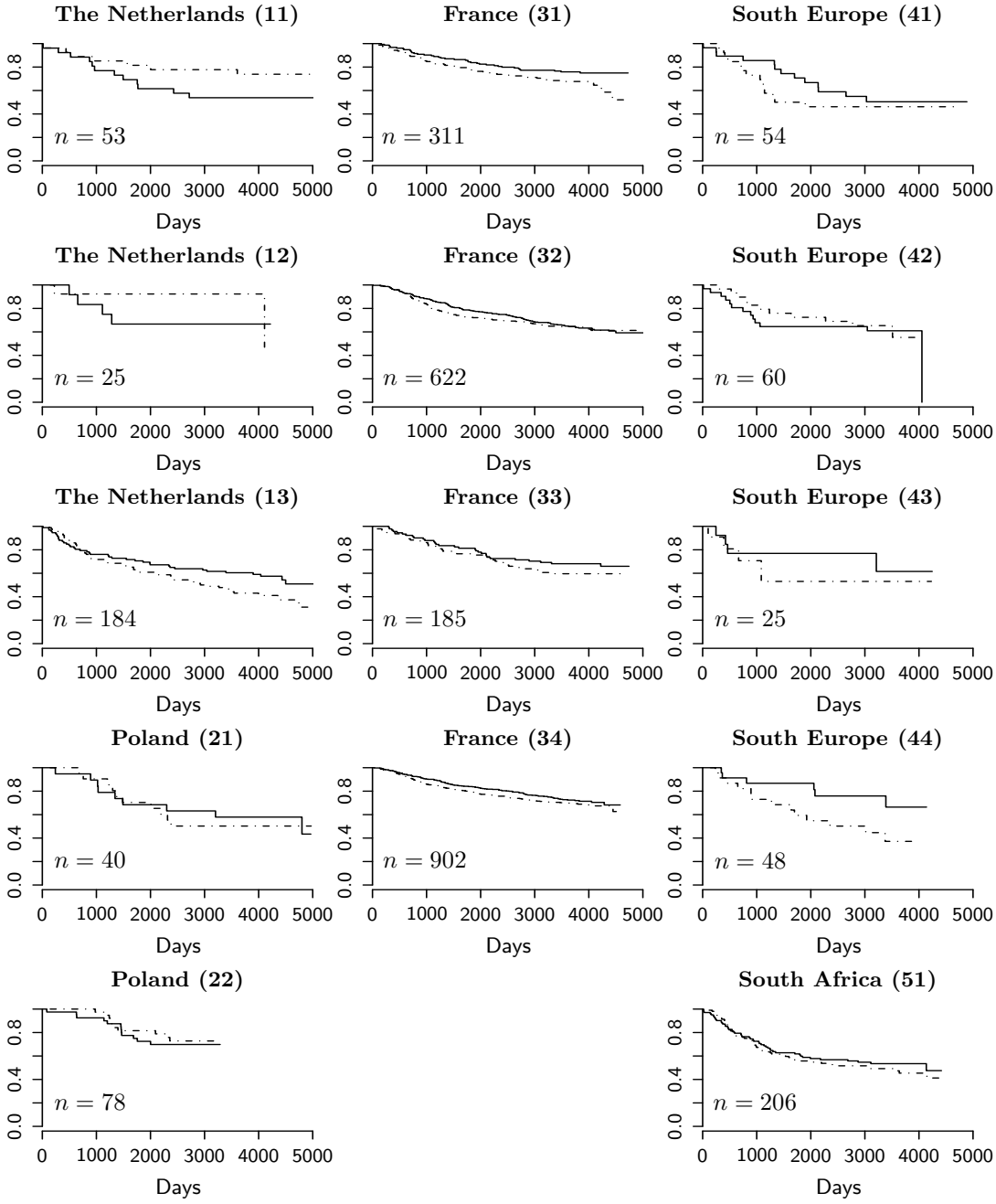

Figure 1. Kaplan-Meier estimates of the DFS time distribution separately for each institution. Solid line: POP FAC arm, dotted-dashed line: control arm.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

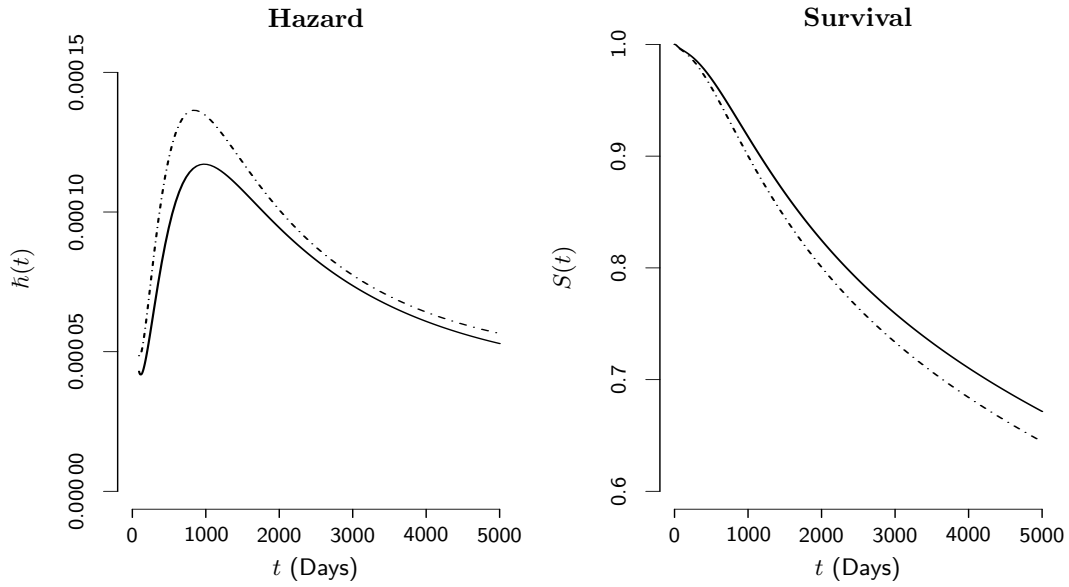

Figure 2. Model with region. Predictive hazard and survival function for the POC FAC arm (solid line) and control arm (dotted-dashed line) and remaining covariates fixed to the reference values.

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls

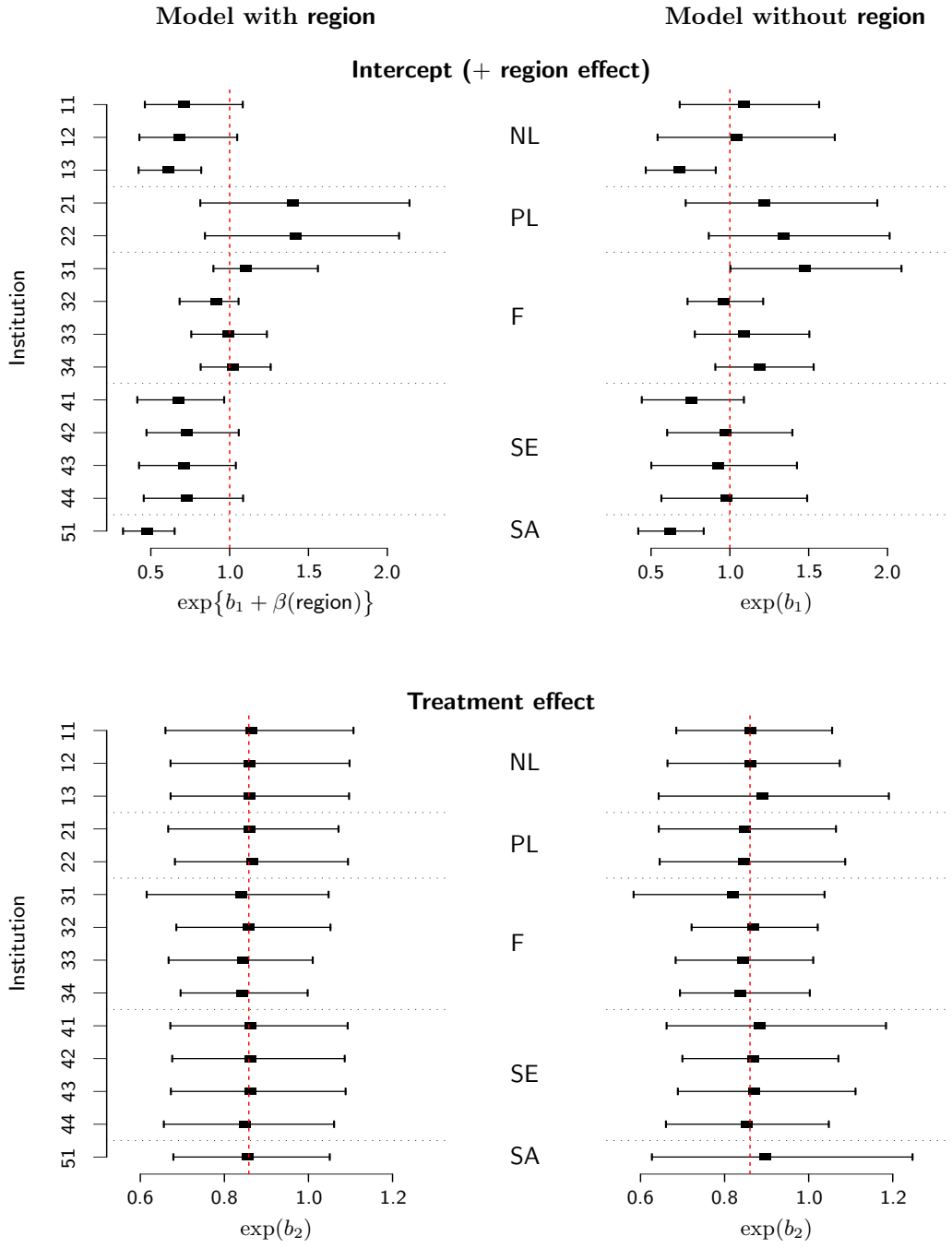

Figure 3. Posterior medians and 95% highest posterior density intervals for center-specific random effects based acceleration factors. Random intercepts in the model with region are further shifted by a corresponding region main effect β(region).

Copyright © 2007 John Wiley & Sons, Ltd.

Prepared using simauth.cls